The non-conditional nature of the Cerf-Adami inequalities and implications for thermodynamics

Ian T. Durham*

Department of Physics, Saint Anselm College, Manchester, NH 03102

(Dated: September 30, 2008)

We show that the Cerf-Adami inequalities do not necessarily depend on conditional entropies nor any reference to Markov chains. While the latter are not explicit in the original form, they are often implied in certain derivations. We also show that these inequalities are intimately related to at least one interpretation of the second law of thermodynamics. The combination of these results provides added insight into why some quantum systems violate the Cerf-Adami inequalities thereby improving our understanding of the quantum-classical boundary. As a result we suggest that the second law may serve as some type of boundary condition on classical knowledge.

I. INTRODUCTION

It has been argued that the laws governing entanglement may well be thermodynamic in nature, or, at the very least, possess thermodynamic corollaries [1, 2]. For example, entanglement has been shown to be necessary in order for the third law of thermodynamics to be consistent with quantum theory [3]. At the heart of entanglement is the notion that the quantum states of two or more objects may be correlated in some way. In 1964, Bell derived an upper bound on the classical strength of these correlations [4, 5] and, since then, numerous experiments have proven that quantum correlations have strengths that exceed this upper bound [6, 7, 8]. Bells derivation and subsequent improvements on his original work have utilized correlation coefficients and expectation values as a measure of entanglement [8, 9]. In 1996, Cerf and Adami introduced the use of entropy as a measure of entanglement and derived an upper bound on the strength of classical correlations using this measure [10, 11]. Entanglement plays a central role in quantum information theory [12] and entropy has long been a measure of information in classical information theory, having been formally introduced by Shannon [13], thus the step taken by Cerf and Adami was a natural one.

The importance of the Cerf-Adami inequalities, as we will call them, lies squarely in the fact that entropy *is* a measure of information. Note that some authors interpret information theoretic entropy in ways that are viewed as more consistent with the thermodynamic (Gibbs-Boltzmann) definition of entropy. For example, Nielsen and Chuang refer to it as the amount of uncertainty that is present in a physical system. However, they note that this makes it ideal for quantifying the resources required to store information [12]. So, however we might look at it, entropy either quantifies the information storage capacity of a system or how much information we

are able to *access* about that system. The Cerf-Adami inequalities utilize a certain type of entropy known as relative entropy that measures information about multiple systems or sub-systems at the same time. For example, suppose we have a tripartite system in which there is a certain amount of information we might have if we knew parts A and B but not C. Conversely there is a certain amount of information we might have if we knew parts B and C but not A, and likewise for A and C but not B. The Cerf-Adami inequalities essentially quantify the relationship of these relative entropies and thus compare the amount of information one might obtain about the system depending on which sub-systems one samples.

The discovery of the Cerf-Adami inequalities proved important for another reason. They pointed to the need for a quantum analogue to the conditional entropy, i.e. a conditional von Neumann entropy. Indeed, it was in [8] that this quantity was *defined*. In addition, they proved to be a generalization of the Braunstein-Caves inequality [14] and thus have proven to be useful in understanding numerous information theoretic problems, e.g. quantum cryptographic protocols [15].

Oddly enough, however, it turns out that the Cerf-Adami inequalities can be derived *without* reference to conditional entropies. The trouble with conditional entropies (and likewise conditional probabilities) stems from the fact that one *could* interpret them as implying a time-like structure, e.g. $H(B|A)$, the entropy of B conditional on knowing A, might be interpreted as implying some knowledge of A must *precede* this knowledge of B. Evidence for this interpretation appears in Cerf and Adami's original paper where they "define the conditional entropy $H(A|B)$ as the entropy of variable A while "knowing", [sic] i.e., having measured, B" [10] which, as worded, implies a previous action. This interpretation is strongly opposed by some [16] thus, a derivation of the Cerf-Adami inequalities that is free of conditional entropies also rids us of at least one debate.

There is also a similar debate concerning Markov chains. While the latter are not *explicitly* utilized in most derivations of the Cerf-Adami inequalities, as we will show they certainly are *implicitly* utilized. Thus, again,

*idurham@anselm.edu

ridding ourselves of the need for Markov chains frees us from another interpretational point, further clearing up the meaning of these inequalities. By then suggesting a potential use for these inequalities in experimentally probing the quantum-classical boundary, we ease the interpretational strain on any possible results.

In order demonstrate all of the nuances inherent in derivations of the Cerf-Adami inequalities, we walk through a derivation based on mutual information that does not include conditional entropies but does include Markov chains. We then present an even simpler derivation that includes *neither* conditional entropies *nor* Markov chains. We encourage the interested reader to compare these to Cerf's and Adami's original paper, in which there is no explicit mention of Markov chains. The first derivation we give below demonstrates why Cerf's and Adami's original derivation *implicitly* relies on Markov chains.

II. THE CERF-ADAMI INEQUALITIES

In information theory it is usual to represent entropy in the binary sense as articulated by Shannon [13],

$$H(X) = -\sum_x p(x) \log p(x) \qquad (1)$$

where the logarithm is taken to be base-two.[28] Suppose we have two systems (or sub-systems) and we wish to measure over the indices x and y. The *joint* entropy measured over these indices is defined as

$$H(X,Y) \equiv -\sum_{x,y} p(x,y) \log p(x,y). \qquad (2)$$

We define the relative entropy [12],

$$H(p(x,y) \parallel p(x)p(y)) \equiv -\sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$$
$$= H(p(x)) + H(p(y)) - H(p(x,y)) \qquad (3)$$

to be a measure of the "offset" of the probability distribution over two indices, x and y, from the probability distributions of the individual indices themselves. As in [12], we define $-0\log 0 \equiv 0$ and $-p(x,y)\log 0 \equiv +\infty$ if $p(x,y) > 0$. Since this represents an offset of the probability distributions, it is zero when these distributions are independent.

The relative entropy can be expressed in a number of ways including as the *mutual* entropy that represents the mutual information of two systems. As such, consider two systems, A and B, that are measured on indices x and y respectively. For convenience (and for ease of transition later) we will dispense with the indices and simply refer to the entropy of the two systems as $H(A)$ and $H(B)$ respectively. We can thus define the mutual entropy as

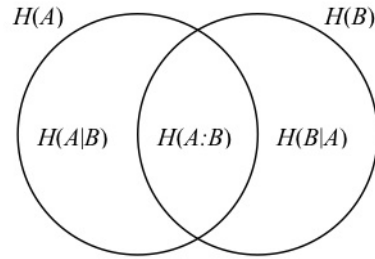$$H(A:B) \equiv H(A) + H(B) - H(A,B). \qquad (4)$$

FIG. 1: In the language of set theory, $H(A{:}B)$ is the intersection of $H(A)$ and $H(B)$. Also note that while it is standard practice to interpret $H(A|B)$ in such a way as to imply some sort of temporal order (see below), in purely set theoretic terms this is not necessary.

See Figure 1 for a visual representation of this and note that $H(A:B) \equiv H(A) \cap H(B)$. Note that equations (3) and (4) imply that

$$H(A:B) = H(p(x,y) \parallel p(x)p(y)) \geq 0 \qquad (5)$$

where the equality holds only when A and B are taken to be *independent* systems measured over independent random variables x and y (that is $p(x,y)=0$).

We define a Markov chain as an *ordered* sequence, $X_1 \rightarrow X_2 \rightarrow \cdots$ of random variables such that X_{n+1} is independent of $X_1, \ldots, X_{n-1}$ given X_n [12]. As such a Markov chain, as defined here, inherently contains the assumption that X_n occurs *before* X_{n+1}. Note that a frequent interpretation extends this such that a series of *measurements* of such variables is also considered a Markov chain [12].

Consider now *three* systems, A, B, and C, each having a corresponding entropy $H(A)$, $H(B)$, and $H(C)$. Suppose we measure the variable X on these systems in such a way that the string of measurements on systems $A \Rightarrow B \Rightarrow C$ is a Markov chain. Then it happens to also be true that the string of measurements on $C \Rightarrow B \Rightarrow A$ is also a Markov chain. In essence there is a temporal order here that is assumed for the measurements of the variable on the systems where each measurement is considered to be independent of any of the others. [29]

With these three systems we may also have $H(B:C)$ and $H(A:C)$ in addition to $H(A:B)$. As Nielsen and Chuang point out, the so called *data processing inequality* supplies an information theoretic description of the conditions under which a Markov chain "loses" information about its earlier values as time progresses. This inequality may be written (in terms of our systems, A, B, and C),

$$\begin{aligned} H(A) \geq H(A:B) \geq H(A:C) \\ H(C) \geq H(C:B) \geq H(C:A) \end{aligned} \qquad (6)$$

where the former is when the chain begins at A and the latter is when the chain begins at C. In other words, if we begin with a certain amount of information about system

A, quantified as $H(A)$, when we proceeded to systems B and C in order, we *lost* information about system A. Essentially $H(A)$ gives us an upper bound on how much *total* information we may possess. Once we gain information about system B, for instance, we must lose *at least* an equivalent amount of information about A such that the total mutual information we have can't exceed our predetermined limit. It is clear from this that there is a time-like progression inherent in this description and that it implies a triangle inequality,

$$H(A:B) + H(B:C) \geq H(A:C) \tag{7}$$

where this is an equality if no information exists about B, i.e. $H(B) = 0$.

If these entropies of the individual systems are normalized (and we are still working with the Markov chain assumption), then the mutual information is symmetric, i.e. $H(A:B) = H(B:A)$, something that should be evident from Figure 1. Equations (6) and (7) together, then, imply

$$H(A:B) + H(B:C) - H(A:C) \leq H(B)$$
$$H(A:B) + H(A:C) - H(B:C) \leq H(A) \tag{8}$$
$$-H(A:B) + H(A:C) + H(B:C) \leq H(C).$$

Further, if the indices that the systems are measured on represent uniform distributions, then $H(A) = H(B) = H(C) = 1$ and we arrive at the Cerf-Adami inequalities,

$$|H(A:B) - H(A:C)| + H(B:C) \leq 1 \tag{9}$$

that have been shown to be in perfect analogy to the usual form of Bell's inequalities and that greatly resemble the Braunstein-Caves inequalities [10, 14].

Why, again, are these inequalities important? First, they serve as a way to define properties of the conditional entropy [10, 11] and put bounds on the sharing of information between systems (parties) thus proving useful for analysis of quantum cryptographic protocols [15]. In addition, violations of these inequalities by certain quantum systems is an indication of the non-separability of those quantum systems [17]. Due to the ubiquity of entropy as a measure in information theory, it makes sense that a set of Bell inequalities based on entropy would be more useful than the original set which are based on correlation coefficients and expectation values.

In any case, we have succeeded here in demonstrating a derivation of the Cerf-Adami inequalities that utilizes Markov chains but not conditional entropies, even though the latter were defined by the original paper that introduced these inequalities. We have also discussed the (potential) temporal nature of Markov chains and conditional entropies. Note that the Shannon entropy is positive definite. Penrose has rigorously shown that any non-decreasing statistical entropy must satisfy two additional constraints: the Markov chain must be *deterministic*, and only the *number* of individual systems can be observed and not their identities [18]. It seems as if the temporal order is a clear requirement, particularly if Markov chains are employed. But what if Markov chains were *not* employed? We shall now show an alternative derivation that does not involve Markov chains and, in fact, makes no reference to any temporal order, either explicitly or implicitly. In fact it avoids an interpretation of the conditional entropy altogether by not using it. Perhaps oddly, we turn to statistical mechanics for this, though the latter is often associated with a temporal order.

III.   AN ALTERNATIVE TREATMENT

There is an alternative definition for the entropy that is commonly used in statistical mechanic and thermodynamic situations that we now introduce. In order to fully articulate it we must first introduce the concept of multiplicity. In most isolated systems there are usually many ways in which the system may configure itself in order to achieve a single macroscopic state, sometimes called a *macrostate*. Each of the ways in which the system may configure itself is usually called a *microstate*. So each macrostate usually consists of several microstates. For example, consider a pair of dice. A roll of 7 would constitute a macrostate. There are six ways in which we might achieve a roll of 7 (assuming the dice are classical) thus there are six microstates associated with the given macrostate. The so-called fundamental assumption of statistical mechanics assumes that all six of these microstates for the roll of 7 are equally probable in the long run. In thermodynamic systems such as a gas an ensemble that has many microstates is often called a microcanonical ensemble. The multiplicity, Ω, of a state is simply the number of microstates for that given macrostate (e.g. it would be six for a roll of 7 on a pair of dice).

In real thermodynamic systems it is often the case that the multiplicity is an enormous number (e.g 10^{120}. Thus, it is often easier and more desirable to logarithmically scale this value. In fact we often define the entropy as

$$H = \frac{S}{k_{\mathrm{B}}} = \ln \Omega \tag{10}$$

where k_{B} is Boltzmann's constant.[30] This is usually called the Boltxmann entropy and is entirely equivalent to equation (1) since the base in both cases is arbitrary (see [19] for a proof of this equivalence[31]). It is quite clear, given the definition of multiplicity, that this definition of entropy is positive definite.

A.   Combinations of systems

Now consider two systems, A and B, with multiplicities $\Omega(A)$ and $\Omega(B)$. Since multiplicity counts microstates, if these systems are combined we would expect the combined systems multiplicity to be a *product* of $\Omega(A)$ and $\Omega(B)$. For example, say system A is a pair of dice showing 7 and system B is a pair of dice showing 8. The multiplicities are six and five respectively. Thus the multiplicity

of a roll of 15 on *four* dice is thirty the multiplicities are multiplicative. The behavior of thermodynamic systems is generally consistent with this idea [18, 19, 20]. So, for example, for systems A and B once they are combined (or considered in unison), the multiplicity of the combination would be

$$\Omega(A,B) = \Omega(A) \cdot \Omega(B) = e^{H(A)} \cdot e^{H(B)} = e^{H(A)+H(B)}. \tag{11}$$

The total entropy, then, is seen to be additive since

$$H(A,B) = H(A) + H(B) = \ln \Omega(A,B). \tag{12}$$

Now this is a very simplified example. In real thermodynamic systems the multiplicity is often a complicated function of volume, number of molecules, temperature, pressure and other thermodynamic quantities. Nonetheless, the additivity of classical entropy is well-established [18, 19, 20] and, in fact, entropy is actually *sub*additive [12], meaning

$$H(A,B) \leq H(A) + H(B) \tag{13}$$

since, in some cases, one can imagine certain microstates might be redundant or might combine (in fact the subadditivity of the entropy is what leads to equation (4) where the equality in (15) holds if they are independent and the mutual entropy is zero).

B. Counting bits

Note that Shannon entropy, while technically unitless, is often measured in bits, i.e. the number of bits of information for a given system. Since we have divided by Boltzmanns constant in (12) in order to make it unitless, there is nothing preventing us from doing the same with the Boltzmann entropy. So suppose we have two systems, A and B, about which we have some information in the form of entropy counted in bits. Suppose further that some of these bits of information actually tell us something about *both* systems simultaneously so these bits count as entropy for both systems. These bits are analogous to a person with dual citizenship who is counted in both his or her countries' censuses. $H(A)$, then, counts *all* the bits that tell us something about system A. Likewise, $H(B)$ counts all the bits that tell us something about system B. The bits that gives us information about *both* systems are technically counted twice, then, since they are included in both $H(A)$ and $H(B)$. By themselves, these bits are labeled $H(A:B)$ since they represent information about *both* systems. The total number of bits we have, that is the joint entropy $H(A,B)$, is

$$H(A,B) = H(A) + H(B) - H(A:B) \tag{14}$$

where we subtract off $H(A:B)$ once so the bits with information about both systems don't get counted twice.

Now consider a third system C. It is trivially true that

$$H(A,B) + H(B,C) \geq H(A,C) \tag{15}$$

where the equality holds if we have no information about system B, i.e. $H(B) = 0$. When we combine this with equation (14), which is simply a rearrangement of equation (4), we find that

$$\begin{aligned} \{H(A) + H(B) - H(A:B)\} \\ + \{H(B) + H(C) - H(B:C)\} \\ \geq H(A) + H(C) - H(A:C). \end{aligned} \tag{16}$$

Reducing and rearranging this produces

$$H(A:B) + H(B:C) - H(A:C) \leq 2H(B). \tag{17}$$

It turns out that we may further narrow this bound. Suppose we have a total of n bits equally distributed among our three systems such that $H(A) = H(B) = H(C) = n/3$ and $H(A) + H(B) + H(C) = n$. Let us assume that it is *not* possible for, say, bits from system A to give information about system B but not the reverse. This means $H(A:B) = H(B:A)$. Given that, suppose *all* of the bits in A also give us information about B. Our previously stated condition requires the reverse to be true. In this case, the total number of bits with "dual citizenship" is $H(A:B)_{\max} = H(A) + H(B) = 2n/3$, or, in the non-maximal case, $H(A:B) \leq 2n/3$. Suppose the same is true for systems B and C. If that were the case, $H(B:C) \leq 2n/3$. Suppose *one* of these two is at a maximum, e.g. $H(A:B)_{\max} = 2n/3$. Since we only have a total of n bits to work with, this limits $H(B:C)$ to a maximum of $n/3$. Adding a third group of shared bits, $H(A:C)$, further reduces this limit. However, by introducing this third group of shared bits we have introduced the possibility of having bits that give us information about *all three* systems. Bits of this sort may be labeled $H(A:B:C)$, but note that we run the risk of counting these bits *three times* since they appear in $H(A)$, $H(B)$, and $H(C)$. Thus

$$H(A) = H(B) = H(C) \leq n/3 \tag{18}$$

and

$$H(A) + H(B) + H(C) - 2H(A:B:C) = n. \tag{19}$$

Suppose $H(A:B)$ is at a maximum. That means that there is no way to distinguish between the bits of system A and those of system B and thus $H(A:B)_{\max} = H(A) = H(B)$. This further implies that $H(A:B:C)$, $H(A:C)$, and $H(B:C)$ all *represent the exact same set of bits* meaning their labels are interchangeable. In other words, in this case,

$$\begin{aligned} H(A:B)_{\max} \Rightarrow \\ H(A:B:C) \equiv H(A:C) \\ \equiv H(B:C) \end{aligned} \tag{20}$$

where we read $\equiv$ as "is identical to" rather than "is defined by" or "is equivalent to" since it means they truly are the same set of bits. Note that it is also true that $H(A) \equiv H(B) = 2n/3$. In any case, these arguments first imply that $H(A : B)_{\max} \leq 2H(B)$ and then, because $H(A : B)_{\max} = H(B)$, they further imply we may drop the factor of 2 as being redundant. This works regardless of which systems we maximally combine since the letters are merely labels for sets of bits. As such we may further narrow the bound on equation (17) and write

$$H(A : B) + H(B : C) - H(A : C) \leq H(B). \qquad (21)$$

Furthermore, when $H(A)$, $H(B)$, and $H(C)$ represent uniform probability distributions, their normalization can be set to unity. Likewise, we could permute the letters depending upon which systems we are comparing. Thus thus may be generalized to equation (9) which are the Cerf-Adami inequalities,

$$|H(A : B) - H(A : C)| + H(B : C) \leq 1. \qquad (9)$$

IV. ENTROPY

As Landau and Lifshitz point out [20], there are inherent difficulties in the interpretation of entropy in terms of the units (i.e. the units are either entirely wrapped up in the multiplicative constant, k_B or the units of bits are assigned to what is technically a unitless quantity such as the Shannon entropy). In fact there are numerous problems inherent in the concept of entropy (see, for example, [18, 19]). As such, the only *uniquely* determined quantities that do not depend on the choice of units are *differences* in entropy, i.e. the changes in entropy brought about by some process [20]. Consider, for example, two systems that are initially separated and then allowed to interact in some manner (for example, two ideal gases separated by a barrier that is later removed). For classical systems, the total entropy of the combined system after mixing is always the same as or greater than it was *before* mixing [19]. In other words, this change in total entropy, often called the *entropy of mixing* [19], is always positive, i.e.,

$$H_{\mathrm{mix}} = \Delta H_{\mathrm{total}} \geq 0 \qquad (22)$$

where the equality holds if the systems are identical (e.g the same type of gas). For example, then, if the systems represent ideal gases and entropy is a method for expressing the probability that a system will be in a given state, the individual entropies provide a method for expressing the probability distributions of the two systems. If the two gases were the same species and otherwise identical prior to mixing, there would be no difference between the two probability distributions and thus no entropy of mixing (technically the multiplicity increases slightly but the factor is negligible and thus it is approximately zero, but always positive regardless).

How might we explain this in terms of bits? Consider two systems of bits, A and B. Note that if there is no mixing, i.e. no bits with 'dual citizenship' (mutual information), then $H(A : B) = 0$. As the number of bits with information about both systems increases, $H(A : B)$ increases. Notice also that this quantity can never be negative *even if we try pulling the systems apart*. It *can decrease*, but it can never be less than zero. Thus the mutual entropy is very similar to the entropy of mixing. It may seem the analogy isnt quite perfect since, in the thermodynamic case the change is in the total entropy while in the information theoretic case the total number of bits appears to remain the same. But there is nothing in the information theoretic case preventing the 'creation' of bits by some other process like noise, for instance. So in thermodynamic systems such as the example of mixing two ideal gases, while $H(A : B)$ increases, we might expect $H(A)$ and $H(B)$ to correspondingly increase which, in fact they do since entropy is a function of volume and by removing the barrier the gases now each have a greater volume through which to spread. Thus we interpret the mutual entropy as a generalization of the entropy of mixing. The mixing process for three systems is described by (14). The entropy of mixing is sometimes interpreted as the work required to mix the systems, but the notion of work is as fraught with problems [20] as that of entropy (perhaps moreso since the problems are largely taxonomic). Either way, we see that changes in the total entropy for isolated systems is never negative (i.e. it never decreases). When pressed for a mathematical statement of the second law of thermodynamics, this is the answer that is frequently given, though usually in a form similar to $\delta S_{\mathrm{isolated}} \geq 0$ where most thermodynamicists use S for entropy. In essence, then, the positivity of the mutual information, argued heuristically a moment ago and clarified in equation (5), is a statement of the second law of thermodynamics. Thus, it is quite clear the the Cerf-Adami inequalities are intimately related to and perhaps even dependent upon the second law of thermodynamics.

As a brief historical note, since the late 1950s, no less than fifteen articles proposing new or revised statements of the second law have appeared *in a single journal*, that being the *American Journal of Physics*. These include a generalized form of the second law of thermodynamics in terms of information that appeared in 1964 [21] and a form *derived from quantum mechanics* that appeared in 1965 [22]! The most recent "new" statement of this law in *Am. J. Phys.* appeared in 1995 [23] while, in 1997, Moore and Schroeder argued in favor of a version (not necessarily new) based on probabilities and multiplicities that is similar to the simple argument we give below examples using coins [24]. Other traditional statements include the well-known version of Kelvin and Planck, that of Clausius, and another known as the Sears-Kestin statement of the second law (see, for example, [25]). Simply put, to this day agreement on a statement of the second law, particularly a mathematically quantifiable one, is strongly

debated. As Partovi recently pointed out, "rarely have so many distinguished physicists written as extensively on a subject while achieving so little consensus" [26].

V. CONSEQUENCES

Thusfar we have demonstrated a relationship between the Cerf-Adami inequalities and the second law of thermodynamics. We have also demonstrated the the aforementioned inequalities can be derived without reference to either Markov chains or conditional entropies. There are several points of significance to this.

A. Conditional entropies and entanglement

Bell-type inequalities are often used experimentally to measure entangled states. In fact certain forms of entropy can be used as a measure of (i.e. to quantify) entanglement. To see this, let us first define the von Neumann as

$$S(\rho) \equiv -\text{tr}(\rho \log \rho) \tag{23}$$

where ρ is a density operator that represents the quantum state of the system. Note that we use S by convention to differentiate it from the classical entropy, H.

The joint entropy of a system with two components, A and B, is

$$S(A, B) \equiv -\text{tr}(\rho^{AB} \log (\rho^{AB})). \tag{24}$$

The *quantum* conditional entropy, that is the quantum entropy of B conditional upon knowing A, is then defined as

$$S(B|A) \equiv S(A, B) - S(A). \tag{25}$$

If $|AB\rangle$ is a pure state of a composite system, $|AB\rangle$ is entangled *if and only if* $S(B|A) < 0$. To quote directly from Cerf's and Adami's original paper, "a violation of an entropic Bell inequality always goes hand in hand with the appearance of a negative conditional entropy" at least from the Venn diagram perspective ([10], p.3). But that is precisely the perspective from which we have derived these inequalities *without* any reference to conditional entropies. Clearly, then, the violation of these inequalities must have its roots elsewhere. Since conditional entropies essentially rescale probability distributions, this seems to imply that violations of Bell inequalities are not necessarily related to conditional probability distributions.

B. Temporal evolution

Deriving the Cerf-Adami inequalities using Markov chains seems to imply there is some temporal order inherent in the inequalities themselves, particularly when considering equation (7) that proceeds directly from equation (6) and the Markov chain assumption. One might be tempted to assume that conditional entropies *also* imply some sort of temporal order, though this is not a universally accepted interpretation of said entropies. Either way it doesn't matter since we have demonstrated a derivation of these inequalities without reference to *either*. This does not necessarily mean that there is no temporal order or evolution inherent in these inequalities. It simply means any such order would have to be associated with something other than the Markov chains and/or conditional entropies. If there *was* such an order inherent in these inequalities, where might it be?

Consider that we have argued that the mutual entropies are a representative statement of the second law of thermodynamics. The second law has often been associated with the arrow of time (though, Partovi has recently demonstrated a reversal of this arrow in macroscopic systems is possible [26]). Thus it might be possible to trace a temporal order to the *mutual* entropy (information). In other words, the temporal order arises from the positivity of the mutual entropy, or, more colloquially, once two systems are mixed it's nearly impossible to perfectly separate them. In fact, one can, of course, speak directly in terms of probability distributions here since that is really what entropies measure. In the Cerf-Adami inequalities, this implies that there is a specific order taken when accessing or measuring the systems.

What does a violation of these inequalities mean then in terms of the second law and the arrow of time? Does it mean that violations of these inequalities implies some violation of causality? Actually, it doesn't and here is why.

1. Reversibility

Ultimately the second law is tied to the idea of reversibility. Consider the following two simple examples.

Example 1 *Suppose we have a single coin. The probability of tossing it and having it land with its head* (H) *showing, is 0.5. The same is, of course, true for its tail* (T). *Suppose we toss it twice and we get* H-T. *So now it is laying with it's tail showing on the back of our hand. Suppose we want to reverse this process, i.e., since it is presently in state* T, *we wish to get it back to state* H. *The probability of doing so is, of course, 0.5. But now, suppose we toss it five times in a row and we get* H-T-H-H-T. *Say we wish to reverse this process, that is we want* T-H-H-T-H. *The probability of accomplishing that is only 0.03125!*

Example 2 *Suppose now we have five coins we wish to flip simultaneously. Say we do so and the result is* H-T-H-H-T. *Say we do so again and the result* T-H-T-T-H. *Suppose we want to reverse this process perfectly, that is, starting with* T-H-T-T-H *showing, we wish to flip the coins such that they return to the state* H-T-H-H-T. *The probability of accomplishing that is also only 0.3125!*

What these two examples show is that a) the further we have progressed through a series of singular random processes, the harder (less probable) it becomes to reverse the series exactly and b) the larger a system is, the harder (less probable) it becomes to reverse a *single* process. This is the heart of the second law of thermodynamics. Many microscopic processes are perfectly reversible, but no *macroscopic* process is (or so we thought until Partovi's recent work [26]). Macroscopic processes may be *approximately* reversible (e.g. the action of opening and closing a door), but it is important to remember that this is only an approximation (open and close the door enough and you'll introduce wear). In this way we might argue that the second law is a strong statement concerning the nature of probabilities and aggregate systems. Indeed, Schroeder has argued this very point [19]. So one way to view the second law and the origin of the arrow of time is as a consequence of constructing macroscopic systems out of many microscopic systems. It's essentially related to the law of large numbers. In our derivations above, we always assumed our distributions were normalized, and thus you'd expect that if the entropies obey the inequality, the probabilities do and the relation of the probabilities to *each other* should not depend on the size of the system. One might then argue that the arrow of time is a result of the inherent tendency of the constituents of the universe to "clump" to form macroscopic objects (we use macroscopic in a broad sense that includes complex molecules, for example). Of course, the discovery of macroscopically entangled and perfectly reversible systems by Partovi adds another element to this argument that we discuss below.

2. An epistemic interpretation

There is another way to look at this, however. We can view relative entropies as providing us with epistemic information about the systems that are involved in the problem. So for instance, $H(A : B)$, the mutual information of systems A and B, could be viewed as a mutual boundary condition on both A and B. So equation (7), for instance, establishes an ordered set of boundary conditions on the knowledge we have (or may obtain) concerning these systems. The second law is sometimes interpreted as being the fact that the universe has an initial boundary condition but not necessarily a final boundary condition. Thus the Cerf-Adami inequalities may be interpreted as placing boundary conditions on the *classical* knowledge we may obtain about a system. In this sense it represents the limit to which our classical knowledge may take us and, beyond which, lies the quantum world whose information is a bit different. Since knowledge may be measured as information in the form of entropies, the generality of the von Neumann entropy arises naturally from this description, i.e. while the von Neu-

mann entropy can *reduce* to a classical entropy in certain situations, it is more general in that it allows for negative probabilities (or, rather, non-seperable density operators). In other words the von Neumann entropy *extends* our knowledge beyond its usual limit. Now, technically this has nothing to do with the second law nor with the thermodynamic arrow of time. As such, it does not necessarily seem immediately true that a violation of the Cerf-Adami inequalities implies the existence of a non-causal process. However, it may be that the second law itself and, by extension the thermodynamic arrow of time, are both simply *boundary conditions on classical knowledge*. In other words, classical information is completely causal while it might be possible that *some* quantum information is not or at least does not have as strict a set of boundary conditions. In other words, classical information, it seems, is entirely governed by initial boundary conditions whereas quantum information seems to be a bit looser in that it could be governed, at least partially, by some unknown *final* boundary condition.

C. The quantum-classical boundary

So, is the microscopic-macroscopic "boundary" discussed in the previous subsection the same as the quantum-classical boundary? Certainly we are not accustomed to thinking of *macroscopic* quantum processes and thus, perhaps, we are inclined to equate the two boundaries. But it may not necessarily be true that they are one and the same. It was Schrödinger's contention that the signature of a quantum system was one that is non-seperable (non-factorable), i.e. entangled (see [27]). Entanglement, as it turns out, has absolutely nothing to do with the size of the system and macroscopic entanglement was recently suggested by Partovi as being associated with so-called ambient correlations [26]. Nor does it necessarily have anything to do with any lengthy string of singular processes such as the first example we gave in the previous section. Understanding the quantum-classical boundary, then, means developing a better grasp of entanglement and *not* necessarily comparing microscopic and macroscopic systems. Would the Cerf-Adami inequalities, then, provide us with a way of probing this boundary? Generally, one might be inclined to think so. But note that there is nothing that *requires* quantum systems violate these inequalities. If we have unentangled states we don't necessarily expect these inequalities to be violated.

What's happening here? Well, the confusion comes from the fact that we refer to quantities such as the von Neumann entropy as being "quantum" when, in reality, the von Neumann entropy is really just a generalized entropy that could be used for *any* system, even a macroscopic, classical one, since, in such a case, it just reduces to the classical entropy. In addition, the language that we refer to as quantum mechanics, might better be thought of as *microscopic* mechanics if we are to take

Schrödinger's view that quantum systems are entangled systems. In a sense, it appears there might be a slight semantic difference that leads to larger conceptual problems. In other words, the Cerf-Adami inequalities do not necessarily get us that much closer to understanding precisely what it is that makes quantum states and systems unique.

Note that, while it might be tempting to consider the uncertainty relations as another sign of "quantumness," the generalized form of those relations as developed by Schrödinger and Robertson are actually purely mathematical relations between certain types of operators and are completely independent of anything "quantum."

[1] M. Horodecki, J. Oppenheim, and R. Horodecki, Phys. Rev. Lett. **89**, 240403 (2002).
[2] J. Oppenheim, M. Horodecki, and R. Horodecki, Phys. Rev. Lett. **90**, 010404 (2003).
[3] M. Wiesniak, V. Vedral, and C. Brukner (2005), quant-ph/0508193v1.
[4] J. Bell, Physics **1**, 195 (1964), reprinted in [5].
[5] J. Bell, *Speakable and Unspeakable in Quantum Mechanics* (Cambridge University Press, Cambridge, 2004), 2nd ed.
[6] A. Aspect, P. Grangier, and G. Roger, Phys. Rev. Lett. **49**, 91 (1982).
[7] A. Aspect, J. Dalibard, and G. Roger, Phys. Rev. Lett. **49**, 1804 (1982).
[8] P. Kwiat and et. al., Phys. Rev. Lett. **75**, 4337 (1995).
[9] P. Kwiat and et. al., Phys. Rev. A **60**, 773 (1999).
[10] N. Cerf and C. Adami, Phys. Rev. A **55**, 3371 (1997).
[11] N. Cerf and C. Adami, Physica D **120**, 62 (1998).
[12] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000).
[13] C. Shannon, Bell System Tech. **27**, 379 (1948).
[14] S. Braunstein and C. Caves, Phys. Rev. Lett. **61**, 662 (1988).
[15] H. Bechmann-Pasquinucci and N. Gisin (1998), quant-ph/9807041.
[16] T. Rudolph (2008), personal communication.
[17] N. Cerf and C. Adami, Phys. Rev. Lett. **79**, 5194 (1997).
[18] O. Penrose, *Foundations of Statistical Mechanics: A deductive treatment* (Dover, New York, 2005), dover ed.
[19] D. Schroeder, *An Introduction to Thermal Physics* (Addison Wesley Longman, Reading, 2000).
[20] L. Landau and E. Lifshitz, *Statistical Physics* (Addison Wesley, Reading, 1958), translated by E. Peierls and R.F. Peierls.
[21] P. Rodd, Am. J. Phys. **32**, 333 (1964).
[22] G. Wannier, Am. J. Phys. **33**, 222 (1965).
[23] A. MacDonald, Am. J. Phys. **63**, 1122 (1995).
[24] T. Moore and D. Schroeder, Am. J. Phys. **65**, 26 (1997).
[25] P.-C. Lu, Am. J. Phys. **50**, 247 (1982).
[26] M. Partovi, Phys. Rev. E **77**, 021110 (2008).
[27] S. Loepp and W. K. Wootters, *Protecting Information: From Classical Error Correction to Quantum Cryptography* (Cambridge University Press, Cambridge, 2006).
[28] Note that the sign conventions employed throughout are consistent with those in [12].
[29] As it turns out, the very nature of Markov chains is somewhat debatable as evidenced by the numerous sometimes contradictory definitions one finds in the literature, e.g. [12, 18, 20].
[30] Note that thermodynamicists usually denote entropy as S rather than H. Information theorists often attribute S to the von Neumann entropy while H is the Shannon entropy. Since the latter is classical we will denote any classical entropy with H for consistency.
[31] Information theorists often use the Shannon entropy and thus base-two. Thermodynamicists tend to use base-ten. In truth, since the logarithm merely scales the multiplicity, it is somewhat arbitrary. The units of entropy in (12) are introduced via Boltzmanns constant and it is not uncommon to actually use the unitless value S/k_{B} instead. Thus, the Shannon entropy, which is technically unitless (though often assigned units of bits by information theorists) can easily be converted to Boltzmann entropy and vice-versa if one converts the multiplicities to probabilities (again, see [19]). Recall that the standard logarithm conversion is $\log_a(x) = \log_b(x)/\log_b(a)$.